Biometrical Journal **58** (2016) 3, zzz–zzz / DOI: 10.1002/bimj.200100000

Hierarchical multivariate mixture generalized linear models for the analysis of spatial data: An application to disease mapping

Mahmoud Torabi[*,1]

[1] *Department of Community Health Sciences, University of Manitoba, S113 Medical Services Building, 750 Bannatyne Ave., Winnipeg, MB, Canada R3E 0W3.*

Received zzz, revised zzz, accepted zzz

Disease mapping of a single disease has been widely studied in the public health set-up. Simultaneous modeling of related diseases can also be a valuable tool both from the epidemiological and from the statistical point of view. In particular, when we have several measurements recorded at each spatial location, we need to consider multivariate models in order to handle the dependence among the multivariate components as well as the spatial dependence between locations. It is then customary to use multivariate spatial models assuming the same distribution through the entire population density. However, in many circumstances, it is a very strong assumption to have the same distribution for all the areas of population density. To overcome this issue, we propose a hierarchical multivariate mixture generalized linear model to simultaneously analyze spatial Normal and non-Normal outcomes. As an application of our proposed approach, esophageal and lung cancer deaths in Minnesota are used to show the outperformance of assuming different distributions for different counties of Minnesota rather than assuming a single distribution for the population density. Performance of the proposed approach is also evaluated through a simulation study.

Key words: Bayesian computation; Exponential family; Hierarchical models; Mixture models; Random effects; Spatial models

1 Introduction

Mapping rates of disease or mortality is essentially a way of describing their spatial distribution over an area. Such distributions are very important for epidemiological and health-policy purposes as mapping rates display the geographic variation in mortality or disease incidence. The main idea behind developments on spatial modeling of disease rates is essentially to model variations in true rates and better separate systematic variability from random noise, a component that usually overshadows crude rate maps. Maps of areal disease incidence and mortality rates are useful tools in determining spatial patterns of the disease for targeting resources. Disease incidence and mortality rates may differ substantially across geographical areas. A reliable estimate of the underlying disease rate is usually provided by borrowing strength from neighboring geographic areas.

For data collected over geographic areas (areal data) such as health authorities, census tracts, and so on, the most commonly used methods to estimate rates are conditionally autoregressive (CAR) specifications (Besag, 1974). The CAR distributions have many applications; including use as the likelihood for the observation in area level models or as the distribution of the random effects in the mean structure in hierarchical models. In the public health set-up, the CAR models are used to study areal patterns of the disease (Torabi, 2012; Torabi and Rosychuk, 2010, 2011, 2012). The CAR models are most appropriate in the univariate case when mapping a single disease is our interest. When we have information for multiple diseases as multivariate areal data, one can obviously use a separate univariate CAR model for each disease. However, since a number of diseases may share the same set of spatial risk factors, or the diseases

*Corresponding author: e-mail: Mahmoud.Torabi@umanitoba.ca, Phone: +1-204-272-3136, Fax: +1-204-789-3905

are related to each other, we may need a multivariate areal model or use a shared latent spatial random risk term to properly analyze this kind of data. This feature enables us to model dependence among the multivariate components while maintaining spatial dependence between areas.

There are several multivariate areal models in the literature. Mardia (1988) considered the theoretical background for multivariate Normal Markov random field (MRF) specification. Billheimer *et al.* (1997) studied a hierarchical statistical model for compositional monitoring data using a multivariate MRF in a state-space setting. A two-fold CAR model for counts of two different diseases over each areal unit was developed by Kim *et al.* (2001). A multi-objective version of the CAR model which allows for flexible modeling of the spatial dependence structure was proposed by Sain and Cressie (2002). Multivariate CAR (MCAR) models for hierarchical modeling based on MRF have been also developed (Carlin and Banerjee, 2003; Gelfand and Vounatsou, 2003). Jin *et al.* (2005) also introduced a flexible class of generalized MCAR (GMCAR) models for areal data using the Bayesian approach. Knorr-Held and Best (2001) proposed a shared component spatial model to study two diseases which are spatially correlated. A common spatial factor model was proposed by Wang and Wall (2003) to study multivariate indicators of cancer risk across counties in Minnesota. Feng and Dean (2012) used a shared latent spatial random risk term to jointly analyze multivariate spatial count and zero-heavy count outcomes. However, the all proposed approaches above assume that all the areas of population density have the same distribution.

Mixture models can provide a flexible framework for modeling the response distribution and can improve model fit. For non-spatial data, there is a well-established literature on mixture models (Mclachlan and Peel, 2000; Fruhwirth-Schnatter, 2006). In the case of spatial setting, several authors have proposed mixture models for point-referenced (geostatistics) data. For example, Gelfand *et al.* (2005) used a Dirichlet process mixture model to examine precipitation measurements at fixed locations in southern France. Kottas and Sanso (2007) extended this approach by allowing the point locations to be random. A similar Poisson point-process mixture model was used to identify cell abundance patterns from fluorescent intensity images of lymphatic tissue (Ji *et al.*, 2009). For multivariate point-referenced data, Reich and Fuentes (2007) proposed a semi-parametric mixture model specified through a stick-breaking process. In the case of areal setting, Green and Richardson (2002) and Lawson and Clark (2002) proposed univariate mixture models for mapping disease risks. Alfo *et al.* (2009) developed a mixture model for multivariate spatial count outcomes using the EM algorithm for the inference. Wall and Liu (2009) also developed a spatial latent class model for multivariate binary data and modeled the latent class indicators using a multinomial probit model with spatially correlated error terms. Recently, Neelon *et al.* (2014) developed a multivariate spatial mixture model for areal data with continuous outcomes to examine areal differences in standardized test scores in North Carolina.

In this paper, we propose a hierarchical multivariate mixture generalized linear model to simultaneously analyze spatial Normal and non-Normal outcomes. As an application of our proposed model, we analyze esophageal and lung cancer deaths in Minnesota. We show that esophageal and lung cancer deaths in Minnesota have a spatially correlated structure (Section 3). We also use mixture models to show, for this dataset, that assuming different distributions for different counties of Minnesota outperforms assuming a single distribution for the population density.

The format of the paper is as follows. In Section 2, we propose a multivariate mixture spatial model in the class of generalized linear mixed model (GLMM). We also study how Bayesian inference can be used to obtain estimates, and also to get prediction for smoothing disease rates (or ratios) over an area. We prove that the corresponding posterior distribution is proper (Theorem 1). Performance of the proposed approach is evaluated using esophageal and lung cancer datasets in Minnesota (Section 3) and by a simulation study (Section 4). Concluding remarks are given in Section 5. The Appendix is devoted to the proof of Theorem 1.

2 Models for multivariate outcomes

2.1 Multivariate spatial generalized linear mixed models

Let y_{ij} be the variable of interest at area i for outcome (or otherwise) $j(j = 1, ..., J; i = 1, ..., n)$. The y_{ij} are assumed to be conditionally independent, given random effects, with exponential family p.d.f.

$$f(y_{ij}|\psi_{ij}, \delta_{ij}) = \exp\{y_{ij}\psi_{ij} - a(\psi_{ij})\}b(y_{ij}), \tag{1}$$

$(j = 1, ..., J; i = 1, ..., n)$. The density (1) is parameterized with respect to the canonical parameters ψ_{ij}, known functions $a(\cdot)$ and $b(\cdot)$. The natural parameters ψ_{ij} are then modeled as

$$\theta_{ij} := h(\psi_{ij}) = q_{ij} + x'_{ij}\beta_j + \phi_{ij} \ (j = 1, ..., J; i = 1, ..., n),$$

where θ_{ij} represents the linear function of fixed and spatial random effects, h is a strictly increasing function, q_{ij} are known constants, $x_{ij}(p \times 1)$ are known design vectors, $\beta_j(p \times 1)$ is a vector of unknown regression coefficients for outcome j, ϕ_{ij} are spatial random effects of area i for outcome j which can be defined as shared spatial component or multivariate CAR models. In particular, we can define $\phi_{ij} = \gamma_j\phi_i$ where γ_j is the factor loading for the shared spatial component on outcome j, with $\gamma_1 = 1$, and ϕ_i is the spatial random effect assumed here to follow a proper CAR distribution (Cressie and Chan, 1989; Stern and Cressie, 1999) to account for the spatially structured correlation in the outcomes. The spatial component

$$(\phi_1, ..., \phi_n)' \sim MVN(0, \Sigma_\phi), \tag{2}$$

$$\Sigma_\phi = \sigma^2(I_n - \lambda C)^{-1}M,$$

where I_n is the identity matrix of dimension n, M is a $n \times n$ diagonal matrix with elements $M_{ii} = 1/g_i; C$ is a $n \times n$ matrix with elements $C_{ik} = 1/g_i$ if areas i and k are adjacent (shown $i \sim k$) and $C_{ik} = 0$ otherwise (also $C_{ii} = 0$), where g_i is the number of areas which are adjacent to area $i; \sigma^2$ is the spatial dispersion parameter; and λ measures the conditional spatial dependence. If $|\lambda| < 1$, $(I_n - \lambda C)$ is then non-singular.

Another way to define spatial random effects is through multivariate CAR models. Let $\Phi_i = (\phi_{i1}, ..., \phi_{iJ})'$, $(i = 1, ..., n)$, be a J-dimensional vector with e.g. ϕ_{i1} as a spatially random variable of the first outcome (e.g., disease) at the ith area. Similar to the univariate CAR models, the unique joint distribution $\Phi = (\Phi_1, ..., \Phi_n)'$ is given by

$$\Phi \sim MVN(0, \Sigma_\Phi), \tag{3}$$

$$\Sigma_\Phi = (\Gamma - C_r)^{-1},$$

where Γ is an $nJ \times nJ$ block diagonal matrix with $J \times J$ diagonal entries Γ_i, C_r is $nJ \times nJ$ with $(C_r)_{ij} = r_iC_{ij}, (C_r)_{ii} = 0$, and r_i is a smoothing parameter at area i. One can choose different Γ and C_r matrices to obtain different MCAR model structure from (3). To obtain a non-singular covariance matrix in (3), we need to make sure that $(\Gamma - C_r)$ is a positive definite and symmetric matrix. However, establishing these conditions may be difficult in the general cases. There have been some work in the literature in that direction. For instance, one can simplify the formulation by assuming that $r_i = \rho I_{J\times J}, (i = 1, ..., n)$, where ρ is again a smoothing parameter, and $\Gamma = D \otimes \Lambda$, where Λ is a $J \times J$ positive definite and symmetric matrix and $D = diag(g_i)$. By replacing these formulas in (3), the corresponding covariance matrix in (3) is positive definite as long as Λ is positive definite and the univariate CAR distribution is valid (Carlin and Banerjee, 2003; Gelfand and Vounatsou, 2003). However, the assumption of a common r_i for all $i(= 1, ..., n)$ may be too strong in some cases. Hence, many MCAR models have been developed for different scenarios of r_i (Kim *et al.*, 2001; Carlin and Banerjee, 2003). One approach is to directly specify the joint distribution for a multivariate spatial process through the specification of simpler conditional

and marginal forms (Jin *et al.*, 2005). For instance, in the case of $J = 2$, let $\eta_1 = (\phi_{11}, ..., \phi_{n1})'$ and $\eta_2 = (\phi_{12}, ..., \phi_{n2})'$ be spatial random variables of two outcomes, one can then have

$$\eta_1 \sim MVN(0, \sigma_1^2(I_n - \lambda_1 C)^{-1}M),$$

and the conditional distribution for $\eta_2|\eta_1$ is

$$\eta_2|\eta_1 \sim MVN(A\eta_1, \sigma_2^2(I_n - \lambda_2 C)^{-1}M),$$

where λ_1 is the smoothing parameter associated with the marginal distribution of η_1, λ_2 is similar for the conditional distribution of $\eta_2|\eta_1$, and σ_1^2 and σ_2^2 are spatial dispersions of η_1 and $\eta_2|\eta_1$, respectively, (Jin *et al.*, 2005). To determine the relationship between η_1 and η_2, it is assumed that the matrix A is $A = \{a_{ik}\}_{i,k=1}^n$ where $a_{ik} = \zeta_0$ if $k = i, a_{ik} = \zeta_1$ if $k \sim i$, and 0 otherwise. Hence, we have $A = \zeta_0 I + \zeta_1 C$ where ζ_0 and ζ_1 are the bridging parameters associating η_{i2} with η_{i1} and $\eta_{k1}(k \neq i)$, respectively.

2.2 Multivariate mixture spatial generalized linear mixed models

A more flexible tool is given in the mixture model approach. The basic idea is that the population under scrutiny may be decomposed into sub-areas with different levels of risk. The parameters describing these levels of risk stem from a discrete parameter distribution $(\pi_1, ..., \pi_L)$. The non-parametric mixture model assumes that these area-specific random effects have a discrete probability distribution taking L values $\theta_{ij1}, ..., \theta_{ijL}$ with probabilities $\pi_{ij1}, ..., \pi_{ijL}$, respectively. Each of these L components of the mixture represents a cluster containing a proportion π_{ijl} (assigning binary variable δ_{ijl} to each cluster l with corresponding probability π_{ijl}) from the population with θ_{ijl} and with the constraint that $\sum_{l=1}^L \pi_{ijl} = 1$. Hence, we propose the following marginal distribution of the outcome y_{ij} which is given by

$$y_{ij} \sim \sum_{l=1}^{L} \pi_{ijl} f(y_{ij}|\psi_{ijl}), \tag{4}$$

$$\theta_{ijl} := h(\psi_{ijl}) = q_{ijl} + x'_{ijl}\beta_{jl} + \phi_{ijl},$$

$$\pi_{ijl} = \frac{e^{x'_{ijl}\upsilon_{jl} + \zeta_{ijl}}}{\sum_{l=1}^L e^{x'_{ijl}\upsilon_{jl} + \zeta_{ijl}}},$$

where we define fixed effects x_{ijl} for each cluster l corresponding to outcome j in area i with the fixed parameters β_{jl}, q_{ijl} are known constants, and ϕ_{ijl} captures spatial random effects for each cluster l and area i corresponding to outcome j. Similar to non-mixture models (Section 2.1), one can define a MCAR model for each cluster l such that $(\phi_{11l}, ..., \phi_{nJl})' \sim MVN(0, \Sigma_{nJ}(\alpha_l))$ depends on model parameters α_l which is $(\lambda_{1l}, \lambda_{2l}, \sigma_{1l}^2, \sigma_{2l}^2, \zeta_{0l}, \zeta_{1l})$ in the case of $J = 2$. Another option is to use factor loading for the shared spatial component (Section 2.1) in the form of $\phi_{ijl} = \gamma_j \phi_{il}$ such that for each cluster l, we have $(\phi_{1l}, ..., \phi_{nl})$ as a proper CAR model (2) with the corresponding parameters $(\lambda_l, \sigma_l^2), (l = 1, ..., L)$. The roles of υ_{jl} and ζ_{ijl} are similar to β_{jl} and ϕ_{ijl}. Note that in model (4), we assume that the number of clusters (L) is known, however, one can consider large number of clusters (say e.g. $L = 10$) and check the performance of different clusters $(L \leq 10)$ using e.g. a Bayesian criterion called deviance information criterion (DIC), (see also the end of Section 2.3), to find out the best choice of L. We will study unknown L with this criterion and other sophisticated approaches in a separate research project (see also the Discussion Section).

2.3 Statistical inference

The multivariate mixture model (4) may be implemented in a Bayesian framework using Markov chain Monte Carlo (MCMC) procedures. The target is to find the posterior distributions of θ and of β given $y = (y_{11}, ..., y_{nJ})$, and in particular, the posterior means, variances and covariances of these distributions. Three special cases are of great practical interest. The first is when the y_{ij} are conditionally independent $Bin(q_{ij}, p_{ij})$ where $\theta_{ij} = \log(p_{ij}/(1-p_{ij}))$ and $a(\psi_{ij}) = q_{ij}\log(1+\exp(\theta_{ij}))$. In the second case, $y_{ij} \sim Poisson(\tau_{ij})$ where $\theta_{ij} = \log(\tau_{ij})$ and $a(\psi_{ij}) = \exp(\theta_{ij})$. In the third case, $y_{ij} \sim N(\theta_{ij}, 1)$ where $a(\psi_{ij}) = \frac{1}{2}\theta_{ij}^2$.

To implement the MCMC, we assume that β_{jl} and parameters involved in the spatial random effects ϕ_{ijl} are mutually independent with $\beta_{jl} \sim$ Uniform $(R^{pL}), (pL < n)$; inverse of variance components have gamma distributions with shape and scale parameters $(\frac{1}{2}c, \frac{1}{2}d)$ with different values of c and d for each variance component; in the case of shared spatial random effects, $\gamma_j \sim$ Uniform (R) and $\lambda_l, (l = 1, ..., L)$, have uniform distributions between $(-1, 1)$.

We now need to check that under the conditions above, the posterior distribution (e.g., in the case of shared spatial random effects for the model (4)) of θ given y is proper. The following theorem provides sufficient conditions to ensure this.

Theorem 1: Assume that $f(y_{ij}|\psi)$ is bounded for all i and j. Suppose also that there exist $y_{i_1 j}, ..., y_{i_m j} (1 < i_1 < ... < i_m \leq n; pJ \leq m \leq n)$ such that

$$\int_{-\infty}^{\infty} \exp\{y_{i_r j}\psi - a(\psi)\}d\psi < \infty,$$

where $r = 1, ..., m$, $\theta = h(\psi), \psi_{ij} = h^{-1}[\sum_{l=1}^{L} \delta_{ijl}(q_{ijl} + x'_{ijl}\beta_{jl} + \gamma_j\phi_{il})]$, and the corresponding design vectors $x_{i_1 jl}, ..., x_{i_m jl}$. Then the joint posterior $\Pr(\theta, \beta, \pi, \gamma, \lambda, \tau|y)$ is proper if $c_l > 0, d_l + n > 0, d_l + m > 0$, and $\tau_l = 1/\sigma_l^2, (l = 1, ..., L)$, where L is finite and known.

Proof of Theorem 1 is given in the Appendix.

The Theorem 1 generalizes the one of Ghosh *et al.* (1999) which considers the univariate spatial model in the non-mixture context with intrinsic CAR (ICAR) for the spatial random effects. Now, the joint posterior distribution (e.g., in the case of shared spatial random effects for the model (4)) can be expressed as

$$\Pr(\theta, \beta, \gamma, \phi, \pi, \lambda, \sigma^2|y) \tag{5}$$

$$\propto f(y|\beta, \gamma, \phi)p(\Phi_1|\lambda_1, \sigma_1^2)...p(\Phi_L|\lambda_L, \sigma_L^2)p(\beta)p(\gamma)p(\pi)p(\lambda_1)...p(\lambda_L)p(\sigma_1^2)...p(\sigma_L^2),$$

where $\sigma^2 = (\sigma_1^2, ..., \sigma_L^2)$. The first term on the right hand side of (5) is the conditional likelihood (1) which is $\exp[\sum_{i=1}^{n}\sum_{j=1}^{J}\{y_{ij}\psi_{ij} - a(\psi_{ij})\}]$. The terms $p(\Phi_1|\lambda_1, \sigma_1^2)...p(\Phi_L|\lambda_L, \sigma_L^2)$ are the distribution of Φ, and the remaining terms are the prior distributions on $(\beta, \gamma, \pi, \lambda_1, ..., \lambda_L, \sigma_1^2, ..., \sigma_L^2)$; flat priors are assigned to β, γ, and λ; priors assigned to the variances of the spatially structured random effects $(\sigma_1^2, ..., \sigma_L^2)$ can be inverse Gamma distributions or their standard deviations can be uniformly distributed, because of the robust properties of this prior (Gelman, 2006); corresponding priors to mixing parameters $(\pi_{ij1}, ..., \pi_{ijL})$ in (4), which are through υ_{jl} and ζ_{ijl}, are assigned similar to β_{jl} and ϕ_{ijl}. Note that we use MCMC with incorporating Metropolis-Hastings to simulate realizations of the posterior distribution. This framework for the analysis was implemented through the freely available software WinBUGS (Lunn *et al.*, 2006).

To compare various models, we use the DIC, defined as $DIC = \overline{D(\Theta)} + p_D$, where $\overline{D(\Theta)}$ is the posterior mean of the deviance with $D(\Theta) = -2\log L(y|\Theta)$ and Θ denotes the collection of parameters in the model (Spiegelhalter *et al.*, 2004). The penalty term p_D is the effective number of model parameters, defined by $p_D = \overline{D(\Theta)} - D(\bar{\Theta})$, where $\bar{\Theta} = E(\Theta|y)$ is the posterior mean of Θ. Models with lower DIC scores are preferred as they achieve a more optimal combination of fit and parsimony.

3 Application to Minnesota lung and esophageal cancers

The data consists of the number of deaths due to lung cancer (y_{i1} observed; E_{i1} expected) and esophageal cancer (y_{i2} observed; E_{i2} expected) in the entire eight year period (1991-1998) at the 87 counties (areas) in Minnesota, USA (Jin *et al.* 2005; Torabi, 2014). The expected numbers of deaths are age-adjusted. Figure 1 shows the standardized mortality ratios (SMRs) of lung cancer (y_{i1}/E_{i1}) and esophageal cancer (y_{i2}/E_{i2}), with some similarity of spatial structure observed in these two cancers. There are also noticeable jumps in the SMR, particularly in the centre and north-centre of Minnesota, where the larger SMRs occur. Figure 1 also shows some substantial differences between the two diseases in terms of their SMRs. For example, counties 13, 35, and 36 have low risk of esophageal cancer, but high risk of lung cancer, while the opposite is true for counties 8, 21, and 46. This supports the choice of a multivariate model which allows different spatial patterns for each disease. A spatial Poisson regression model was used as these cancer deaths are assumed to be rare enough relative to the population in each county. The model is then given by

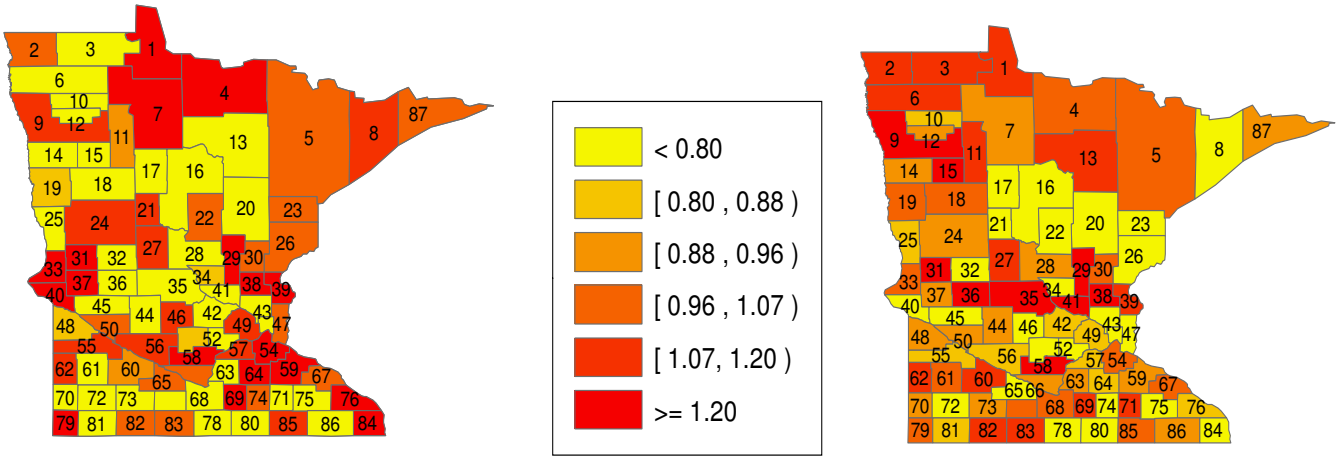
$$y_{ij} \sim \sum_{l=1}^{2} \pi_{ijl} Poisson(\psi_{ijl}),$$

$$\theta_{ijl} := \log(\psi_{ijl}) = \log(E_{ij}) + x_i'\beta_{jl} + \phi_{ijl}, (i = 1, ..., 87; j = 1, 2; l = 1, 2),$$

$$\pi_{ijl} = \frac{e^{x_i'\upsilon_{jl}+\zeta_{ijl}}}{\sum_{l=1}^{L} e^{x_i'\upsilon_{jl}+\zeta_{ijl}}},$$

where π_{ij2} denotes the weight for the second mixture component, $\zeta_{ijl} \overset{ind.}{\sim} N(0, \sigma_\zeta^2)$, x_i' with x_{1i} is proportion of *males* in the ith county, and x_{2i} as ethnicity covariate which is proportion of *black* in the ith county. It seems that the risk factors of sex and ethnicity are the same for the both lung cancer and esophageal cancer cases, so in our study we consider $\beta_{jl} = \beta_l$ (Jemal *et al.*, 2008); noting that it is well known that smoking is a risk factor for lung cancer, however, we did not have access to this covariate in our dataset. We also consider two different structures (factor loading and multivariate model) for the spatial random effects ϕ_{ijl}. By end of this analysis, we also evaluate that $L = 2$ seems to be enough for our dataset. As indicated in this paper, we assume that the number of clusters (L) is known.

We considered $c/2 = 1$ and $d/2 = 10$ ~~$c = 0.001$ and $d = 1$~~ as shape and scale parameters of priors in the inverse gamma for our variance components. We ran 3 initially dispersed chains for $150,000$ iterations each, discarding the first $50,000$ as burn-in. To reduce autocorrelation, we retained every 50 iteration. To monitor the convergence of the model parameters, we used several diagnostic methods implemented in the Bayesian output analysis program (Smith, 2007), a freely available package created for R (R Development Core Team, 2014). In particular, we evaluated descriptive diagnostic tests such as the autocorrelation of generated samples of model parameters from the posterior distribution and convergence diagnostic tests such as Brooks, Gelman, and Rubin tests (Gelman and Rubin, 1992; Brooksand and Gelman, 1998) and Heidelberger and Welch test (Heidelberger and Welch, 1983). None of these tests indicated non-convergence of the model parameters. We also investigated the choice of priors through a sensitivity study (e.g., using uniform distribution for standard deviations rather than inverse gamma distribution for variance components) for our data analysis and found our results are robust against priors.



(a) (b)

Figure 1 Standardized mortality ratios of esophageal cancer (a) and lung cancer (b) in Minnesota.

Model diagnostics indicated efficient mixing and rapid convergence of the chains. In particular, the post-burn-in trace plots (not shown here) indicated that the chains overlapped substantially. To evaluate the performance of our model, we also compared our proposed model (in both cases of spatial random effects: factor loading and multivariate model) to three sub-models: (1) a mixture model with separate spatial random effects ($M3$); (2) a non-mixture model with separate spatial random effects ($M6$); and (3) a non-mixture model with shared ($M4$) and multivariate ($M5$) spatial random effects. It is worth mentioning that we also tried $\log(\theta_{ijl}) = \log(E_{ij}) + x_i'\beta_{jl} + \gamma_{jl}\phi_i$, however, we had a convergence issue.

Table 1 provides the model comparison results for the various models, noting that ϕ_{ijl} in model $M2$ and ϕ_{ij} in model $M5$ indicate the multivariate spatial random effects. The mixture model with factor loading of shared spatial random effects ($M1$) outperformed other sub-models, noting that the model with shared spatial random effects ($M1$) also performed better than the corresponding model with multivariate spatial random effects ($M2$). Overall, the proposed mixture spatial model had the best performance, suggesting that incorporating the shared spatial random effects into the mixing weights provided a modest additional benefit relative to other sub-models.

Table 2 presents the posterior means and 95% credible intervals (CrIs) for the proposed model parameters. The results suggested that there are two distinct mixing components, or "latent sub-populations" of cancer patients. Sub-population 1 contained an estimated 58% of the overall population, and was characterized by comparatively high proportions of other ethnicities (compared to the *black*) and same size of males and females ($\beta_{11} = -0.04, \beta_{21} = -3.42$). Sub-population 2 contained the remaining 42% of the overall population and was associated with higher proportions of females with slightly larger group of other ethnicity compared to the black group ($\beta_{12} = -10.15, \beta_{22} = -0.40$). In addition, the model accounted

Table 1 Model comparison statistics for analysis of the lung and esophageal cancer mortality data.

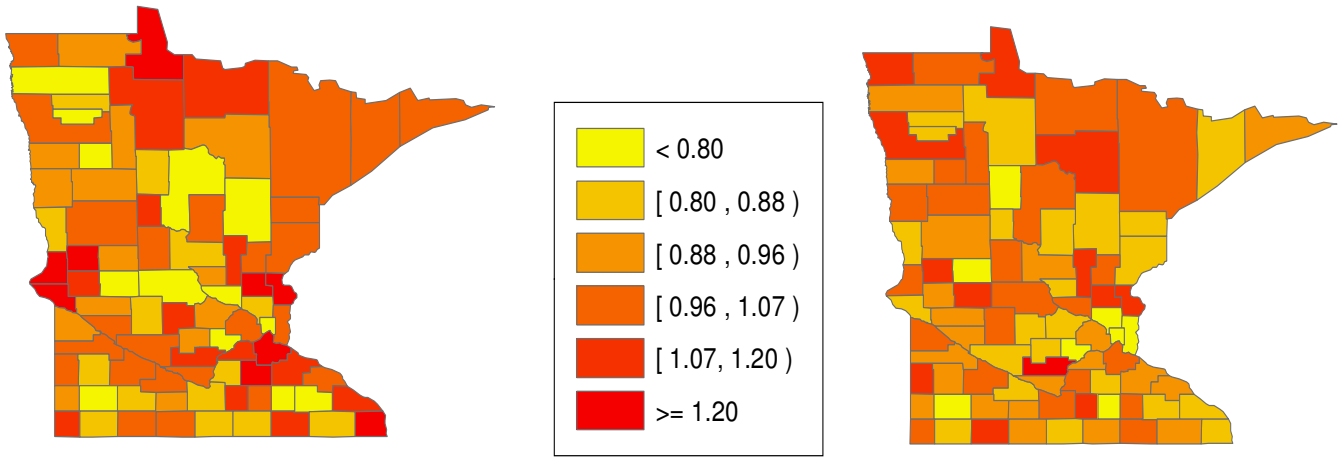
Model	$\overline{D(\theta)}$	pD	DIC
$M1 : \log(\psi_{ijl}) = \log(E_{ij}) + x_i'\beta_{jl} + \gamma_j\phi_{il}$	1101	46	1147
$M2 : \log(\psi_{ijl}) = \log(E_{ij}) + x_i'\beta_{jl} + \phi_{ijl}$	1137	45	1182
$M3 : \log(\psi_{ijl}) = \log(E_{ij}) + x_i'\beta_{jl} + \phi_{il}$	1141	42	1183
$M4 : \log(\psi_{ij}) = \log(E_{ij}) + x_i'\beta_j + \gamma_j\phi_i$	1147	40	1187
$M5 : \log(\psi_{ij}) = \log(E_{ij}) + x_i'\beta_j + \phi_{ij}$	1153	38	1191
$M6 : \log(\psi_{ij}) = \log(E_{ij}) + x_i'\beta_j + \phi_i$	1160	36	1196

Table 2 Posterior means and 95% credible intervals (CrIs) for the proposed model M1.

Mixture component	Parameter	Posterior Mean	95% CrI
1	β_{11}	-0.04	$(-0.40, 0.36)$
	β_{21}	-3.42	$(-5.39, -1.23)$
	λ_1	0.96	$(0.87, 0.99)$
	σ_1^2	0.04	$(0.02, 0.07)$
2	β_{12}	-10.15	$(-18.90, -2.32)$
	β_{22}	-0.40	$(-2.30, 1.60)$
	λ_2	0.50	$(0.40, 0.60)$
	σ_2^2	0.14	$(0.06, 0.22)$
	γ_2	0.85	$(0.68, 1.02)$

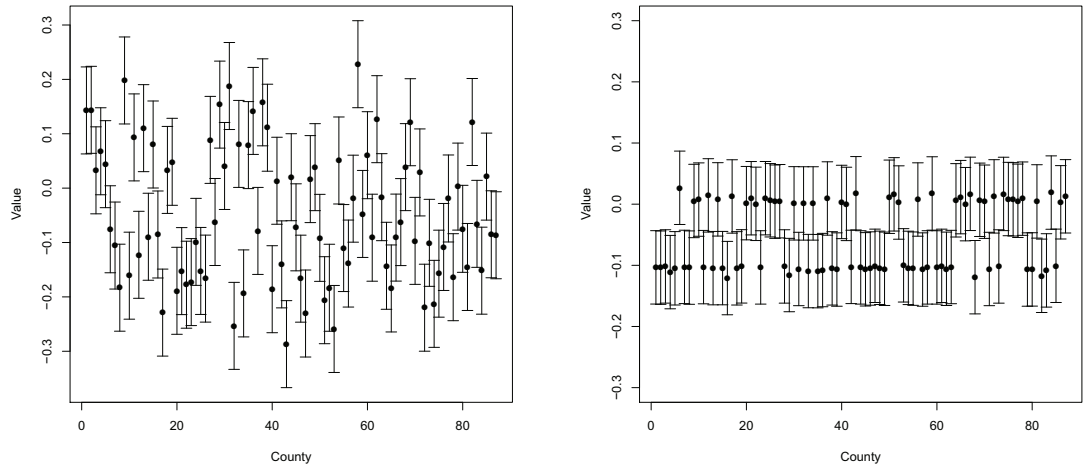
for spatial association in both sub-populations which were relatively high ($\lambda_1 = 0.96$ and $\lambda_2 = 0.50$). The factor loading parameter γ was estimated as $0.85(0.68, 1.02)$, indicating that the both cancers share a common and possibly identical spatial structure, with γ substantially above zero and with $\gamma_2 = 1$ not rejected. The posterior expected SMRs of the proposed model is given in Figure 2. The common spatial component resembles the SMR maps for lung and esophageal cancers (Figure 1). The mixture model appears to have maintained some areas with jumps (centre and north-centre) for both cancers while displaying increased smoothing in other parts of Minnesota. Based on our available data, we can conclude that other ethnicity rather than blacks and females are two distinct groups who are most at risk to die due to lung and esophageal cancers; these two groups also have different spatial patterns of deaths due to lung and esophageal cancers over the counties in Minnesota. In particular, the focus may be in the counties with the higher ratios of death due to lung and esophageal cancers, compared to the average population, for further investigations and possible interventions.

The posterior mean estimates, along with the 95% CrIs of the common latent spatially structured effect for two sub-populations are shown in Figure 3. It seems that the residual terms appear to be more or less flat, confirming the dominance of a strong underlying spatial structure shared between lung and esophageal cancers. To examine whether any residual spatial structure has been left for each of the proposed model, we used Moran's I (Moran, 1950; Feng and Dean, 2012; Neelon *et al.*, 2014; Cliff and Ord, 1981; Cressie, 1993) tests on the county residuals from the model. In particular, let $\hat{y}_{ij}$ denotes the predicated disease counts for cancer $j(= 1, 2)$. Moran's I statistic is given by $I_j = \frac{e_j' D e_j}{e_j' e_j}$, where $e_j' = (e_{1j}, ..., e_{nj})$ with $e_{ij} = (y_{ij} - \hat{y}_{ij})/\sqrt{var(\hat{y}_{ij})}$. The posterior mean (95% CrI) of I_1 and I_2 for lung and esophageal cancers are $0.015(-0.049, 0.068)$ and $0.047(-0.023, 0.129)$, respectively, suggesting that the residuals have no significant spatial correlation.



(a) (b)

Figure 2 Posterior expected SMRs of esophageal (a) and lung (b) cancers in Minnesota.



(a) (b)

Figure 3 Point estimates and 95% posterior credible intervals for the spatial random effects of first (a) and second (b) sub-population; analysis of the lung and esophageal cancer dataset.

Table 3 Mean values of the model parameters estimates and corresponding 95% coverage probability, the variance of the estimated parameters and mean values of the estimated variances for spatial Poisson mixture model based on $1,500$ simulated datasets.

				Variance	
Mixture component	Parameter	Mean	95% coverage probability	HB	Simulated
1 (% 55)	β_{11} =-0.01	-0.005	0.940	0.004	0.003
	β_{21} =0.10	0.093	0.943	0.002	0.002
	λ_1 =0.40	0.568	0.972	0.067	0.011
	σ_1^2 =0.12	0.116	0.945	0.017	0.008
2 (% 45)	β_{12} =-0.07	-0.067	0.941	0.007	0.005
	β_{22} =0.04	0.042	0.958	0.005	0.004
	λ_2 =0.40	0.550	0.973	0.069	0.009
	σ_2^2 =0.12	0.116	0.931	0.022	0.016
	γ_2 =0.82	0.82	0.948	0.00002	0.00002

4 Simulation study

We also conduct a simulation study to evaluate performance of the spatial Poisson (non-Normal outcomes) mixture model using a scenario similar to our Minnesota lung and esophageal cancer dataset. More specifically, data are generated from the following model:

$$y_{ij} \sim \sum_{l=1}^{2} \pi_{ijl} Poisson(\psi_{ijl}), \tag{6}$$

$$\log(\psi_{ijl}) = \log(N_i) + \beta_{jl} + \gamma_j \phi_{il}, (i = 1, ..., 87; j = 1, 2; l = 1, 2),$$

where $\gamma_1 = 1$, and with parameters $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \gamma_2, \lambda_1, \lambda_2, \sigma_1^2, \sigma_2^2, \pi_{ij1}, \pi_{ij2})$ listed in Table 3. The neighborhood structure and the population sizes (US Census 2010 dataset) are from the Minnesota dataset. Estimates are obtained using $1,500$ datasets generated from the spatial Poisson mixture model (6). The mean values of the model parameters estimates, the corresponding 95% coverage probability, the variance of the estimated parameters, and mean values of the variances are presented in Table 3. It seems that the estimates of model parameters are reasonably unbiased, and their variances are also estimated well with comparing the variances with the corresponding simulated values. Overall, it seems that the spatial Poisson mixture model provides good point estimates and corresponding variances for this data analysis.

5 Discussion

Disease mapping studies that consider only one disease have been widely used. However, simultaneous modeling of multiple diseases is also very important provided we have measurements recorded at each spatial area which are believed to be related. In addition, we may have different underlying distributions for different areas of population density. To address this issue, we have proposed a multivariate mixture spatial model in the class of GLMM to simultaneously analyze spatial Normal and non-Normal outcomes. We used the Bayesian approach for the estimation of model parameters, and also prediction of the smoothing disease ratios (rates) over an area. A motivation for our work was esophageal and lung cancer deaths in Minnesota. Based on our available data from Minnesota, we concluded that other ethnicity rather than blacks and females are two distinct groups who are most at risk to die due to lung and esophageal cancers. In particular, we found that some counties in the centre and north-centre of Minnesota have higher ratios of death due to lung and esophageal cancers compared to the average population. These findings may

represent real changes in those counties or different distributions of important covariates that are unmeasured and unadjusted for in our modeling. Further investigation may be warranted to explore these findings.

Our approach is very general in the context of SGLMM using CAR models with shared components or multivariate spatial random effects. In this paper, we assumed that the number of clusters is known, however, this may not be an appropriate assumption in some applications. We have planned to study the SGLMM for areal data assuming the number of clusters is unknown. As a natural extension of our model, we have also planned to study a hierarchical multivariate mixture generalized linear model to simultaneously analyze spatio-temporal Normal and non-Normal outcomes.

Acknowledgements I would like to thank the Editor, an Associate Editor, and a referee for constructive comments and suggestions, which led to the improvement of the paper. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflict of Interest
The author has declared no conflict of interest.

Appendix

Proof of Theorem 1: Using one-to-one transformation, one can write $z_{il} = \phi_{il} - \phi_{nl}, (i = 1, ..., n-1), z_l = (z_{1l}, ..., z_{(n-1)l})'$, and then

$$\Pr(\theta, \beta, \pi, z, \delta, \gamma, \lambda, \tau | y) \propto \exp[\sum_{i=1}^{n}\sum_{j=1}^{J}\{y_{ij}\psi_{ij} - a(\psi_{ij})\}]$$

$$\times \Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}[\tau_l^{(\frac{1}{2})n}]^{\delta_{ijl}}\Pi_{k=1,k\neq i}^{n}\exp[\frac{-1}{2}\delta_{ijl}\tau_l\lambda_l(z_{il}-z_{kl})^2 C_{ik}]$$

$$\times \Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}[\tau_l^{(\frac{d_l}{2})-1}]^{\delta_{ijl}}\exp[\frac{-1}{2}\delta_{ijl}c_l\tau_l]\pi_{ijl}^{\delta_{ijl}}$$

$$\times \Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}\pi_{ijl},$$

where $z_{nl} = 0, \beta = (\beta_{jl})_{j,l}^{J,L}, \pi = (\pi_{ijl})_{i,j,l}^{m,J,L}, z = (z_l)_l^L, \delta = (\delta_{ijl})_{i,j,l}^{m,J,L}, \gamma = (\gamma_j)_j^J, \lambda = (\lambda_l)_l^L, \tau = (\tau_l)_l^L$. Without loss of generality, we assume that $i_r = r (r = 1, ..., m)$ and let $\theta_* = (\theta_{il}, ..., \theta_{ml})'$ where $\theta_{il} = (\theta_{i1l}, ..., \theta_{iJl})'$. Now, integrating with respect to $(\theta_{(m+1)l}, ..., \theta_{nl})$, the joint posterior of $\theta_*, \beta, \pi, z, \delta, \gamma, \lambda, \tau$ given y is

$$\Pr(\theta_*, \beta, \pi, z, \delta, \gamma, \lambda, \tau | y) \propto \exp[\sum_{i=1}^{m}\sum_{j=1}^{J}\{y_{ij}\psi_{ij} - a(\psi_{ij})\}]$$

$$\times \Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}[\tau_l^{(\frac{1}{2})n}]^{\delta_{ijl}}\Pi_{k=1,k\neq i}^{n}\exp[\frac{-1}{2}\delta_{ijl}\tau_l\lambda_l(z_{il}-z_{kl})^2 C_{ik}]$$

$$\times \Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}[\tau_l^{(\frac{d_l}{2})-1}]^{\delta_{ijl}}\exp[\frac{-1}{2}\delta_{ijl}c_l\tau_l]\pi_{ijl}^{\delta_{ijl}}$$

$$\times \Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}\pi_{ijl},$$

then,

$$\Pr(\theta_*, \beta, \pi, z, \delta, \gamma, \lambda, \tau | y) \propto \exp[\sum_{i=1}^{m}\sum_{j=1}^{J}\{y_{ij}\psi_{ij} - a(\psi_{ij})\}]$$

$$\times\Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}[\tau_l^{\frac{1}{2}(n+d_l)-1}]^{\delta_{ijl}}\Pi_{k=1,k\neq i}^{n}\exp[\frac{-1}{2}\delta_{ijl}\tau_l\{c_l+\lambda_l(z_{il}-z_{kl})^2C_{ik}\}]$$

$$\times[\Pi_{i=1}^{n}\Pi_{j=1}^{J}\pi_{ijl}^{\delta_{ijl}}]\times[\Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}\pi_{ijl}].$$

Now, integrating with respect to τ gives

$$\Pr(\theta,\beta,\pi,z,\delta,\gamma,\lambda|y)$$

$$\leq K\exp[\sum_{i=1}^{m}\sum_{j=1}^{J}\{y_{ij}\psi_{ij}-a(\psi_{ij})\}]$$

$$\times\Pi_{i=1}^{n}\Pi_{k=1,k\neq i}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}[c_l+\lambda_l(z_{il}-z_{kl})^2C_{ik}]^{\frac{-\delta_{ijl}}{2}(n+d_l)}$$

$$\times[\Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}\pi_{ijl}^{\delta_{ijl}}]\times[\Pi_{i=1}^{n}\Pi_{j=1}^{J}\Pi_{l=1}^{L}\pi_{ijl}],$$

where $K(>0)$ is a generic constant which does not depend on θ_* and z. We also know that β,γ, and λ have uniform distributions, and δ_{ijl} given other parameters follows a multinomial distribution $(1,\pi_{ij1},...,\pi_{ijL})$, with υ_{jl} and ζ_{ijl} have similar behavior as β_{jl} and ϕ_{ijl}, respectively. In addition, $z_{nl}=0$ and $\sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n}\sum_{l=1}^{L}\lambda_l(z_{il}-z_{kl})^2C_{ik}$ involves only $(n-1)L$ variables $z_{1l},...,z_{(n-1)l}$, $(l=1,...,L)$. Thus, integrating with respect to z, and using the structure of a multivariate t-distribution, it follows that

$$\Pr(\theta|y)\leq K\exp[\sum_{i=1}^{m}\sum_{j=1}^{J}\{y_{ij}\psi_{ij}-a(\psi_{ij})\}],$$

which completes the proof of Theorem 1.

References

[1] Alfo, M., Nieddu, L. and Vicari, D. (2009). Finite mixture models for mapping spatially dependent disease counts. *Biometrical Journal* **51**, 84–97.

[2] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society, Series B* **36**,192–236.

[3] Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H-W. and Silkey, M. (1997). Natural variability of benthic species composition in the Delaware Bay. *Environmental and Ecological Statistics* **4**, 95–115.

[4] Brooksand, S. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.

[5] Carlin, B.P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). *In Bayesian Statistics 7*, JM Bernardo, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AF Smith, and M West (eds.), 45–63. Oxford University Press.

[6] Cliff, A.D. and Ord, J.K. (1981). *Spatial Processes: Models and Applications*, Pion: London.

[7] Cressie, N.A.C. (1993). *Statistics for Spatial Data* (revised ed.), Wiley: New York.

[8] Cressie, N.A. and Chan, N.H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association* **84**, 393–401.

[9] Feng, C.X. and Dean, C.B. (2012). Joint analysis of multivariate spatial count and zero-heavy count outcomes using common spatial factor models. *Environmetrics* **23**, 493–508.

[10] Fruhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*, Berlin: Springer-Verlag.

[11] Gelfand, A.E., Kottas, A. and MacEachern, S.N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.

[12] Gelfand, A.E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**, 11–25.

[13] Gelman, A.E. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

[14] Gelman, A.E. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**,457–511.

[15] Ghosh, M., Natarajan, K., Waller, L.A. and Kim, D. (1999). Hierarchical Bayes GLMs for the analysis of spatial data: An application to disease mapping. *Journal of Statistical Planning and Inference* **75**, 305–318.

[16] Green, P.J. and Richardson, S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association* **97**,1055–1070.

[17] Heidelberger, P. and Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operational Research* **31**,1109–1144.

[18] Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., Thun, M.J. (2008). Cancer Statistics, 2008. *CA: A Cancer Journal for Clinicians* **58**, 71–96.

[19] Ji, C., Merl, D., Kepler, T.B. and West, M. (2009). Spatial mixture modelling for unobserved point processes: Examples in immunofluorescence histology. *Bayesian Analysis* **4**, 297–316.

[20] Jin, X., Carlin, B.P. and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61**, 950–961.

[21] Kim, H., Sun, D. and Tsutakawa, R.K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a two-fold conditional autoregressive model. *Journal of the American Statistical Association* **96**, 1506–1521.

[22] Knorr-Held, L. and Best, N.G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society, Series A* **164**, 73–85.

[23] Kottas, A. and Sanso, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference* **137**, 3151–3163.

[24] Lawson, A.B. and Clark, A. (2002). Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine* **21**,359–370.

[25] Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2006). WinBUGS a Bayesian modelling framework: concepts, structures, and extensibility. *Statistics and Computing* **10**, 325–337.

[26] Mardia, K.V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis* **24**, 265–284.

[27] Mclachlan, G. and Peel, D. (2000). *Finite Mixture Models*, New York: John Wiley & Sons.

[28] Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23.

[29] Neelon, B., Gelfand, A.E. and Miranda, M.L. (2014). A Multivariate spatial mixture model for areal data: Examining regional differences in standardized test scores. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **63**, 737–761.

[30] R Development Core Team (2014). A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Available at http://www.R-project.org.

[31] Reich, B.J. and Fuentes, M. A. (2007). Multivariate semi-parametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics* **1**, 249–264.

[32] Sain, S.R. and Cressie, N. (2002). Multivariate lattice models for spatial environmental data. *Proceedings of the American Statistical Association, Statistics and the Environment Section* [CD-ROM], 2820–2825. Alexandria, Virginia: American Statistical Association.

[33] Smith, B.J. (2007). *BOA User Manual (version 1.1.7)*, Department of Biostatistics, College of Public Health, University of Iowa, Ames.

[34] Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2004). *WinBUGS version 1.4 User Manual*, MRC Biostatistics unit, Institute of Public Health, London.

[35] Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 795–809.

[36] Stern, H.S. and Cressie, N.A. (1999). Inference for extremes in disease mapping. *In Disease mapping and risk assessment for public health*. A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J-F. Viel and R. Bertollini (eds.), Chichester: Wiley, 63–84.

[37] Torabi, M. (2012). Hierarchical Bayes estimation of spatial statistics for rates. *Journal of Statistical Planning and Inference* **142**, 358–365.

[38] Torabi, M. (2014). Likelihood inference in spatial generalized linear mixed models with multivariate CAR models for areal data. *Journal of Spatial Statistics* **10**, 12–26.

[39] Torabi, M and Rosychuk, R.J. (2010). Spatio-temporal modelling of disease mapping of rates. *The Canadian Journal of Statistics* **38**,698–715.

[40] Torabi, M and Rosychuk, R.J. (2011). Spatio-temporal modeling using spline for disease mapping: Analysis of childhood cancer trends. *Journal of Applied Statistics* **38**, 1769–1781.

[41] Torabi, M and Rosychuk, R.J. (2012). Hierarchical Bayesian spatiotemporal analysis of childhood cancer trends. *Journal of Geographical Analysis* **44**, 109–120.

[42] Wall, M.M. and Liu, X. (2009). Spatial latent class analysis model for spatially distributed multivariate binary data. *Computational Statistics & Data Analysis* **53**, 3057–3069.

[43] Wang, F. and Wall, M. (2003). Generalized common spatial factor model. *Biostatistics* **4**, 569–582.

